

BASEMENT





ON THE RELIABILITY AND PREDICTIVE VALIDITY OF
PURCHASE INTENTION MEASURES†

Manohar U. Kalwani*
Alvin J. Silk**

SSM W.P. # 1255-81

July 1981

THE MARKETING CENTER

Massachusetts Institute of Technology
Alfred P. Sloan School of Management
50 Memorial Drive
Cambridge, Massachusetts 02139
(617) 253-6615

ON THE RELIABILITY AND PREDICTIVE VALIDITY OF
PURCHASE INTENTION MEASURES†

Manohar U. Kalwani*
Alvin J. Silk**

SSM W.P. # 1255-81

July 1981

* Krannert Graduate School of Management, Purdue University

** Sloan School of Management, Massachusetts Institute of Technology

† We are indebted to Harry Davis for supplying us with data from an unpublished study he conducted which we used in the work reported here. Thanks are due to Michael Sider for his assistance in data analysis. Portions of this paper were presented at the TIMS/ORSA Meetings, Colorado Springs, October 1980; and the TIMS Market Measurement Conference, New York University, April 1981.

ABSTRACT

This paper reports some further analyses and applications of Morrison's model of the predictive relationship between measures of intentions and subsequent purchasing behavior. A review of published studies bearing on the threats to predictive validity of intention scales represented in Morrison's model is presented. Findings from a test-retest study of intention ratings for concept stimuli are shown to be consistent with the levels of reliability expected under the model's assumptions of beta binomial distributed scores. Evidence of the predictive validity of intention measures is found in a re-analysis of several sets of relevant data but a different form of predictive relationship is shown to hold for generic durable goods as compared to branded packaged goods. Whereas a linear relationship is supported in the case of durable goods, the presence of a threshold phenomenon in the branded packaged goods data suggests the use of a piecewise linear model. There is reason to believe that the sources of systematic error present in intentions ratings are different for these two types of purchases.

INTRODUCTION

Purchase intention is one of a very small set of variables that find routine application in consumer research investigations undertaken for a variety of different purposes (e.g., new product concept and copy tests; segmentation and tracking studies) and covering a broad range of products and services.¹ The collection of purchase intention data has become fairly standardized in the sense that one or the other of two basic types of instruments appear to be employed in much of this work, either a five point intention scale (definitely will buy = 5; definitely will not buy = 1) or the eleven point purchase probability scale developed by Juster (1966). Despite widespread utilization of these scales, reservations and criticisms have been expressed frequently about this reliance on intention measures with the issue of predictive ability being the paramount concern. Do intentions expressed at a particular point in time relate to future purchase events?

The published evidence available on this matter is much more abundant for purchasing behavior at the level of generic types of durable goods than for the purchasing of specific brands within product categories--either durable or non-durable. Numerous studies conducted by economists have found statistically significant relationships between prior intentions and subsequent purchasing of generic durables at both the individual household and aggregate levels but the forecasting value of intentions data has been disappointing and controversial (McNeil 1974; Adams 1974; and Juster 1974). However, turning to the matter of intentions to buy particular brands, one finds only bits and pieces of data in the marketing research literature relevant to the question of predictive validity. On a few occasions, intention studies

have been followed up with efforts to obtain reports of subsequent purchasing behavior from the same sample of respondents. While positive associations between intentions and purchases have generally been observed in such data (Gormley 1974; Laroche and Howard 1980; Penny, Hunt, and Twyman 1972; Rothman 1964; Smith, Davenport, and Parker 1963; Tauber 1975; Taylor, Houlahan, and Gabriel 1975; Twyman 1973; and Wells 1961), the strength of the relationship uncovered in these analyses has not been viewed as sufficiently marked and consistent to allay the basic concern about the predictive validity of intention scales noted above and shared by many in the marketing research community.

Another body of research has been accumulating on the psychometric properties of intention scales. Published studies have reported evidence bearing on the vulnerability of these intention measures to response style biases (Clancy and Garsen, 1970), their ability to discriminate among alternative brands or product concepts (Gold and Salkind, 1974; Haley and Case, 1979; and Ptacek and Ross, 1980), and their sensitivity to the effects of advertising exposure (Axelrod, 1968). The presence of substantial components of random and/or systematic errors in intention measures will, of course, tend to affect adversely both their discriminant and predictive abilities. What past research in this area has lacked is an integrating framework which would allow assessments of the discriminant and predictive validity of intention scales to take explicit account of the major sources of fallibility present in such measurements.

Recently, Morrison (1979) has made an important contribution to enhancing understanding of the quality of intention measures and to improving their practical usefulness. Morrison has proposed a model of the linkage between intentions and purchasing which, among other things, represents several threats to the predictive validity of intentions measures and lends itself to statis-

tical estimation and testing. Thus for the first time, a formal model has been specified for intention scaling similar in spirit to the psychometric models which are the foundation of psychological measurement.

The purpose of the present paper is to report some further analyses and empirical applications of Morrison's model. We begin with a summary of the model, emphasizing its psychometric interpretation and correcting a minor error in one of the original equations. In the second section of the paper, we review the evidence available relating to two threats to the predictive validity of intention measures represented in Morrison's model: random measurement error or unreliability and systematic error or bias. Some empirical estimates of the reliability of a seven point intention scale are reported from a small scale test-retest study of new product concepts. An examination of the evidence on bias draws attention to possible differences in the properties of intention scales according to whether they are applied to generic durable goods, a brand of frequently purchased consumer products, or concept test stimuli.

The third section of the paper discusses the results obtained by applying Morrison's model to two different classes of intention-purchase behavior data. One body of data pertains to the purchasing of generic types of durable goods in the United Kingdom. The intention measure is the same eleven point purchase probability scale used in the well-known Juster (1966) study which Morrison reanalyzed. The second set of data examined involves brand purchasing of packaged goods and here the instrument used to measure intentions was either a five point or a seven point scale. Thus we investigate the applicability of Morrison's model to a broader range of situations than those he examined for illustrative purposes in his original paper. Maximum likelihood methods are used for parameter estimation and likelihood ratio

tests are conducted to compare the goodness-of-fit of two alternative specifications of the relationship between the measures of intentions and purchasing: (a) the linear relation used by Morrison, and (b) a piecewise linear model suggested by empirical findings reported by Haley and Case (1979) and Laroche and Howard (1980) with reference to packaged goods and by the "top box" scoring rules frequently used by marketing research practitioners in analyzing five point intention scale data (Gold and Salkind 1974; and Gruber 1970). We conclude with a discussion of the implications of our findings for further testing of the model and its applications.

MODEL STRUCTURE

Figure 1 summarizes the overall structure of Morrison's model. The principal aim is to model the "predictive" relationship between the two observables: an individual consumer's intentions measured at a particular point in time (I_x) and his/her probability of purchasing (P_x) as manifested in the subsequent purchase events that occur and are reported after some time interval (Δ). Intentions are measured with error (ϵ) and hence expressed intentions (I_x) are an imperfect indication of true intention (I_t) which may change over the time interval and have a different unobserved true value (I_t') due to a host of possible influences (e.g., unexpected change in income, later consumption experience, etc.) all of which are summarized in a single change parameter (ρ). Intentions may also be subject to systematic sources of error (response style biases mentioned previously, the effects on promotion and the like) and hence Morrison proposes to adjust for such biases via a parameter, b . Thus the true predictive validity of intentions may be obscured in the observed relation with purchasing due to the presence of three types of threats: (a) random measurement error in expressed intentions, (b) instability or change, and (c) bias in true intentions.

=====

INSERT FIGURE 1 HERE

=====

Morrison models random measurement error at the individual level by assuming a respondent's expressed intention (I_x) is a random variable binomially distributed with parameters I_t (true intention) and n , where $n+1$ is the number of points or alternative response categories presented in the intention scale. True intentions are assumed to be beta distributed in the population and then the aggregate unconditional distribution of expressed intentions across consumers is beta binomial. A key feature of the beta binomial formulation stressed by Morrison is that it leads to a linear relationship between the expected value of true intention and the expressed or measured intention. Below we discuss some further interpretations of the beta binomial assumption and investigate its implications regarding scale reliability.

The Reliability of Beta Binomial Intention Scales

Respondents are asked to express their purchase intentions by checking one of the $n+1$ labelled response categories which the instrument provides. The response alternatives are ordered to represent a monotonic increasing (or decreasing) level of intention and responses are assigned an arbitrary integer score $x = 0, 1, 2, \dots, n$ which reflects this ordering.

Modeling a respondent's behavior in selecting a response alternative as if it were the outcome of a binomial process is a strong assumption which Morrison (1979, p.67) points out is "untestable" but useful in that it provides a "reasonably specific representation" of the conventional measurement model where the observed score is taken as a sum of a true value and a random error component. Morrison hypothesizes that "an individual with true intention I_t responds 1 or 0 in an independent fashion to each point on the scale with probabilities I_t and $1-I_t$, respectively," and hence the observed

intention (x) is "the sum of these 0, 1 responses" (p.67). A somewhat different way to conceive of the respondent's behavior is to presume that he/she regards the specified response alternatives as forming an interval scale (where $I_x = x/n$) and then selects the particular category whose value is closest to his/her particular true intention I_t . Uncertainty in choosing the appropriate discrete response category, momentary fluctuations in attention, and various other kinds of transitory factors could lead a respondent to check different points on the scale if the task were repeated and the resulting frequency of response is postulated to be binomial. Such a line of reasoning is akin to Thurstone's notion of a "discriminal dispersion" used in formulating his laws of comparative and categorical judgement which assume a normal distribution of responses (Torgerson 1958, pp. 156-158). A useful way to illuminate the meaning of the binomial assumption here is to examine the curves describing the probability of endorsing the various response categories as a function of the individual's underlying true intention. The family of such curves, known as the trace lines or operating characteristics (Torgerson 1958, p. 362), are shown in Figure 2 for a five point scale. Thus, a respondent with true intention (I_t) of .75 would be most likely to check 4 on a five point scale but would have a non-zero probability of choosing any of the other points as shown in Figure 2.

=====
INSERT FIGURE 2 HERE
=====

To obtain an expression of the reliability of an $n+1$ point intention scale consistent with Morrison's beta binomial assumptions, we begin with the usual measurement model:

$$I_x = I_t + \epsilon, \quad (1)$$

where I_x and I_t represent observed and true intentions, respectively, and ϵ is the random measurement error. For fixed true score, I_t , the observed score (I_x) is distributed binomial with:

$$E[I_x | I_t] = I_t, \quad (2)$$

and,

$$\text{Var}[I_x | I_t] = \frac{I_t(1-I_t)}{n}, \quad (3)$$

where the expectation and variance operators are over the respondent population. Equations (2) and (3) imply that, for a fixed true score I_t , the measurement error is unbiased, that is:

$$E[\epsilon | I_t] = 0, \quad (4)$$

and

$$\text{Var}[\epsilon | I_t] = \frac{I_t(1-I_t)}{n}. \quad (5)$$

It is worth noting that the conditional distribution of the measurement error, ϵ , is not independent of the true score, I_t . This is to be expected since the model contains bounded observed scores ($0 \leq I_x \leq 1$) and unbiased measurement errors.

Recall that reliability of a scale is defined as the proportion of the variance in observed scores accounted for by the variance in true scores (Lord and Novick 1968, p. 61). For the model set forth in equation (1), the reliability of the intention scale is given by:

$$R^* = \frac{\text{Var}[I_t]}{\text{Var}[I_x]}. \quad (6)$$

The variance of I_x which follows the beta-binomial distribution is:

$$\text{Var}[I_x] = \text{Var}[I_t] + E \left[\text{Var}[\epsilon | I_t] \right]. \quad (7)$$

The first term on the right hand side of (7) represents the variance of the conditional means and the second term is the mean of the conditional variances. Using equation (5) and carrying out the expectation and variance operations we find the two right-hand side terms in equation (7) simplify to (see Appendix A):

$$\text{Var}[I_t] = \alpha\beta/(\alpha+\beta)^2(\alpha+\beta+1) , \quad (8)$$

$$\left[E \text{ Var}[\epsilon | I_t] \right] = \alpha\beta/n(\alpha+\beta)(\alpha+\beta+1) . \quad (9)$$

Adding the two variance terms in equations (8) and (9) yields:

$$\text{Var}[I_x] = \alpha\beta(\alpha+\beta+n)/n(\alpha+\beta)^2(\alpha+\beta+1) . \quad (10)$$

Therefore, the reliability of the beta binomial distributed intention scale is given by:

$$\begin{aligned} R^* &= \frac{\alpha\beta/(\alpha+\beta)^2(\alpha+\beta+1)}{\alpha\beta(\alpha+\beta+n)/n(\alpha+\beta)^2(\alpha+\beta+1)} \\ &= \left(\frac{n}{\alpha+\beta+n} \right) \end{aligned} \quad (11)$$

The reliability expression for the beta binomial intention scale in equation (11) has several interesting features.² First, reliability of the beta binomial distributed intention scale will differ across stimuli according to variations in the sum, $\alpha+\beta$. Reliability will increase as the term $\alpha+\beta$ approaches zero. This is intuitively appealing since when the term $\alpha+\beta \rightarrow 0$, the beta distribution has virtually all of its mass concentrated near 0 or 1. Then for most people, the measurement error in equation (5) would be near zero and the intention scale would be highly reliable. At the other extreme, the scale will become completely unreliable as the term $\alpha+\beta$ becomes large. As Morrison (1979, p.67) notes, this is consistent with the fact that the beta binomial distribution becomes a "spike" at its mean $\alpha/(\alpha+\beta)$ when the term $\alpha+\beta \rightarrow \infty$. Under such circumstances, all respondents would have the same true intention so that the measurement error present would dominate the variance in the true scores.

A second important property is that reliability of a beta binomial intention scale is expected to improve when the number of scale points is increased. Thus, given $\alpha = \beta = 0.5$, the reliability of a five point scale ($n = 4$) is expected to increase from $4/5 = 0.80$ to $10/11 = 0.91$ when the number of scale points is raised to eleven ($n = 10$). This relation, (11) corresponds to the result found in classical test theory, the Spearman-Brown formula, which indicates that the reliability of

a composite score increases monotonically as a test is lengthened by adding more items which are homogeneous with respect to the original set in the sense of preserving the pattern of inter-item correlations (Lord and Novick 1968, pp. 112-224). Thus, in Morrison's model of a single or one item rating scale, the number of response alternatives or levels it contains plays an analogous role to that of the number of items in the standard psychometric model of a composite score derived from multiple items.

Relating Purchase Intentions and Purchase Behavior

Citing the work of Keats and Lord (1962), Morrison (1979, p. 67) points out that a beta binomial distributed scale will yield a linear relationship between the expected value of true intention, I_t , and stated intention, I_X . He notes that this linear relation is given by:

$$E[I_t | I_X] = \frac{\alpha}{\alpha + \beta + 1} + \left(\frac{1}{\alpha + \beta + 1} \right) I_X \quad (12)$$

Our derivation of this relationship (see Appendix B) leads to a slightly different result:

$$E[I_t | I_X] = \left(\frac{\alpha}{\alpha + \beta + n} \right) + \left(\frac{n}{\alpha + \beta + n} \right) I_X \quad (13)$$

Thus we find that the equation (12) given by Morrison (1979, p. 67) contains an error and should be replaced by the corrected version (13).

Morrison (1979, p. 66) suggests a modification in the beta binomial model to account for changes in true intentions of individual respondents during the time frame of interest due to exogenous factors such as unexpected changes in income. His proposed modification does not affect the linear nature of the relationship between the expected value of I_X and I_t . Basically, the modification allows a change to occur in an individual's true intention with probability, p , and given such a change, the individual is assigned a new true intention I_t' , randomly drawn from the original beta density function. The new relationship between the expected value of I_t' and I_X is given by (see Morrison 1979, p. 68):

$$\begin{aligned} E[I_t' | I_x] &= \frac{\rho\alpha}{\alpha+\beta} + \frac{(1-\rho)\alpha}{\alpha+\beta+n} + \frac{(1-\rho)n}{\alpha+\beta+n} I_x, \\ &= \rho E[I_t] + (1-\rho)E[I_t | I_x], \end{aligned} \quad (14)$$

where $E[I_t' | I_x]$ denotes the expected value of revised true intention. This modification in the beta binomial model leads to a reduction in the slope of the linear relationship between the expected value of I_t and I_x .

As mentioned earlier, intention measures are susceptible to systematic sources of error due to response style tendencies, promotional effects, changes in the economy, etc. In Morrison's model, bias introduced in $E[I_t' | I_x]$ by these systematic errors is assumed to be constant across all respondents. The overall bias for the sample of respondents in equation (14) is:

$$b = \bar{I}_x - \pi, \quad (15)$$

where $\bar{I}_x$ is the mean of the observed intention scale values and π is the overall proportion of respondents who subsequently purchase. Morrison (1979, p. 68) adjusts for systematic errors by subtracting the overall bias b from each $E[I_t' | I_x]$:

$$\begin{aligned} P_x &= E[I_t' | I_x] - b \\ &= \frac{\rho\alpha}{\alpha+\beta} + \frac{(1-\rho)\alpha}{\alpha+\beta+n} + \frac{(1-\rho)n}{\alpha+\beta+n} I_x - b \\ &= \frac{\rho\alpha}{\alpha+\beta} + \frac{(1-\rho)}{\alpha+\beta+n} + (1-\rho)R^* I_x - b, \end{aligned} \quad (16)$$

where P_x is the probability of purchasing given expressed intention, I_x and R^* is the reliability index defined in (11). Hence we see from the final model (16) that the predictive validity of stated intentions, I_x 's depends upon their stability (ρ), reliability (R^*) and biasedness (b).

RESEARCH ON RANDOM AND SYSTEMATIC ERRORS IN INTENTIONS SCALES

In this section we review evidence bearing on two properties of intention scales which represent potential threats to their predictive validity specified in Morrison's model. The first is reliability, which relates to the random error component present in intention ratings--a subject which appears to have been largely ignored in past research. We present some empirical estimates of reliability and compare them to the values expected under the assumption that the ratings follow a beta binomial distribution. Bias is the second threat discussed below along with the attendant problem of discriminant ability. Consideration is given to how systematic errors may vary according to whether intentions are measured with reference to the purchasing of generic durable goods or branded packaged goods. Finally the issue of how many response categories should be used in an intention scale is examined.

Reliability

A review of the marketing research literature indicates that surprisingly little attention has been given to the reliability of intention scales. The only published reliability estimate for an intention scale the present authors were able to uncover is that found in Ptacek and Ross (1979). The estimates were based on data collected in a concept test wherein each respondent provided an intention rating for each of seven new product stimuli. The ANOVA model (Ebel 1951; Winer 1971, pp. 283-293) used to estimate the variance components makes the highly restrictive assumption that all respondents had the same true intention score for a given concept. Within concept variations in ratings were attributed to systematic between-respondent mean differences in response style biases plus random measurement error and hence no allowance was made for individual differences in true intention. The reliability

coefficients reported used between concept mean differences rather than the customary between respondent differences as the true score variance estimator. It can be shown (Winer 1971, p. 291) that the reliability of a rating computed in this manner is approximately equal to the average intercorrelation between ratings of the seven concepts given by pairs of respondents. Thus, the small values of the reliability coefficients reported by Ptacek and Ross (.110 for the 5 point intention scale and .176 for the eleven point Juster scale) are probably best interpreted as an indication of respondent heterogeneity rather than as evidence of the low reliability of the individual respondents intention ratings.

The only other data found in the literature bearing on the issue of scale reliability arise from situations where both the eleven point purchase probability scale and the five point intention scale have been administered to the same sample of respondents with reference to the same products or stimuli. Such a study may be viewed as a test-retest situation with two different measures of the same trait, purchasing intentions. If the beta binomial model holds separately for each intention scale and if respondents' true intentions are constant across administrations of the two scales and thus are beta distributed with the same parameters (α and β), then the distributions of the two sets of observed ratings will be beta binomial but differ due to differences in the number of scale points used in each scale (5 vs. 11). It is readily shown that the product-moment correlation between pairs of such intention ratings (X_1, X_2) scales is given by:

$$r(X_1, X_2) = \left(\frac{n_1 n_2}{(n_1 + \alpha + \beta)(n_2 + \alpha + \beta)} \right)^{1/2}$$

$$= (R_1^* R_2^*)^{1/2}$$

Thus the product-moment correlation between intentions ratings obtained on, say, five and eleven point scales from the same population of respondents with reference

to the same product or stimuli is equal to the geometric mean of the reliabilities of the individual scales, R_1^* and R_2^* . Assuming $r(X_1X_2) > 0$, then $r^2(X_1X_2)$ is a lower bound for the two reliabilities since $0 \leq R^* \leq 1$:

$$R_1^* \geq r^2(X_1X_2) \leq R_2^* .$$

Juster (1966, p. 677) presents a cross-tabulation of responses to a six month buying intention question involving five response categories asked with reference to automobiles against responses to his eleven category purchase probability scale obtained "a few days" later from the same sample of 447 consumers. The product-moment correlation between these two scales was found by the present authors to be 0.576 which would imply a lower bound of 0.332 on the reliabilities of the eleven point and five point intentions measures used by Juster under the assumptions of the beta binomial model.

Another piece of evidence indicating that the five point intention scale and the eleven point Juster probability scale are highly intercorrelated is found in Haley and Case (1979) factor analysis of thirteen widely used rating scales administered among a sample of 630 women with reference to the six leading brands in six categories of packaged goods. The rotated loadings on the main "evaluative" factor were very high for both intention scales: .88 for the five point scale and .86 for the eleven point Juster scale. The between scale correlation indicated by these factor loadings is .76 and implies a lower bound of .57 for their respective reliabilities.

Harry Davis generously made available to us data from an unpublished study he conducted of the test-retest reliability of a seven-point intention scale administered in a concept test setting. The stimuli rated were descriptions of 20 diverse new products and services that spanned purchases involving major as well as small financial outlays. The same convenience sample of 63 adult

male business executives rated each of the 20 concepts on two occasions separated by a two week interval. Table 1 lists the concepts and the computed values of their respective test-retest correlation coefficients.

 INSERT TABLE 1 HERE

As discussed in Silk (1977), a necessary condition for a test-retest correlation coefficient to serve as an interpretable reliability coefficient in the sense of (6) is that the variances of the observed test and retest scores be equal. This condition appeared to hold here as indicated by the results of separate tests for the equality of two correlated variances carried out for each of the concepts. For only one of the 20 concepts could the null hypothesis of homogeneous test and retest variances be rejected at even the .20 level (2 tail test). However, this overall conclusion must be qualified inasmuch as beta binomial distributed intention scores do not satisfy the distributional assumption (bivariate normality) underlying this test (Snedecor and Cochran 1967, pp. 195-197).

Accepting that the condition of stable test and retest variances was not seriously violated, we may compare the observed levels of reliability indicated by the test-retest correlations ($\hat{r}_{12}$) with those expected for beta binomial distributed intention scales as given by (11) above. For a seven point intention scale ($n = 6$) we see from (11) that the expected reliability (R^*) will vary inversely with the magnitude of $\alpha + \beta$ according to $R^* = \frac{6}{\alpha + \beta + 6}$. A plot of this relationship appears as the solid line in Figure 3 for values of $\alpha + \beta$ ranging 0 to 7. The distributions of intention ratings from the first test were used to estimate the beta binomial parameters α and β for each concept by means of the method of maximum likelihood (Kalwani (1980)). These estimates are shown in Table 1 along with the expected values of the reliability index computed by substituting the estimates,

$\hat{\alpha} + \hat{\beta}$, in (11) for $n=6$. As indicated by the values of the chi square statistics reported in Table 1, the beta binomial model generally provided a good approximation to the distributions of the intention ratings for these concepts obtained from the initial test occasion. Comparing the computed and critical values of the chi square statistics shown in Table 1, we find that the null hypothesis that the intention ratings are distributed binomial could be rejected at the .05 level for 6 of the 20 concepts and at the .01 level for only 3 of the 20 concepts. In Figure 3, the 20 pairs of points corresponding to the test-retest correlation and the sum of the estimates of the two beta binomial parameters for each concept are also plotted.

 INSERT FIGURE 3 HERE

The concepts are listed in Table in descending order according to the magnitude of the sum, $\hat{\alpha} + \hat{\beta}$. In light of the great diversity of the concepts, the considerable variation on the estimated values of α and β is reassuring in that it provides a preliminary indication that the respondents were at least somewhat discriminating in rating the different concepts. For half the concepts, the intentions ratings tend to be skewed toward the high end of the scale while in the other half they tend to be skewed in the opposite direction. However, in only about a third of the cases is the skewness very marked in either direction.

Except for the air conditioner concept, in all cases the sum, $\alpha + \beta$, exceeds unity -- the range of values of the quantity being from .920 to 6.494. Accordingly, for these data the reliability coefficient, $R^* = \frac{6}{\alpha + \beta + 6}$ would be expected to vary from .867 to .480 and in Table 1 we find the observed values of the test-retest correlations encompass this range, varying from .840 to .506. More critically, we see in Figure 3 that the test-retest correlations tend to decrease as $\hat{\alpha} + \hat{\beta}$ increases in a manner that appears to conform generally with the expected form of the relationship indicated by $R^* = \frac{6}{\alpha + \beta + 6}$. The median of the absolute differences

between the value of the expected reliability coefficient (R^*) and the test-retest correlation ($\hat{r}_{12}$) for the 20 concepts is .077 (on the 0-1 scale), with the value of $\hat{r}_{12}$ exceeding that of R^* for 9 of the 20 concepts. In light of the fact that the same respondents rated all 20 concepts, it is noteworthy that the scatter of the observed test-retest correlations around the expected reliability curve shown in Figure 3 does not appear to follow any systematic pattern.

To sum up, then, these results indicate that the reliability of a seven point intention scale behaves in a manner that is generally consistent with Morrison's assumption of an underlying measurement model consisting of a beta distributed true score and a binomial distributed error component. Clearly the scope of the evidence presently available to assess this assumption is limited and further studies of reliability are needed, especially for scales with different numbers of scale points and for applications of them to both branded packaged and durable goods.

Bias and Discriminant Ability

The literature on intention scales shows much more concern with bias or systematic error than with random measurement error and reliability. Evidence relating to systematic errors in intention scales is available from two quite different bodies of research.³ One group of studies has investigated the biases in intention measures which appear when they are related to purchase data. In the other set of studies the frequent failure of respondents to discriminate among alternatives in their intentions ratings for a set of stimuli has been attributed to the presence of a persistent and stable component of systematic error.

1. Intentions and Purchases

In the area of purchases of generic types of household durables, considerable evidence exists documenting the systematic error that appears when responses to intention questions or the Juster probability scale obtained at one point in time are compared cross-sectionally to reports of subsequent purchase events for the same sample of households. An excellent review and discussion on this work appears in McNeil (1974, 1975), Adams (1974), and Juster (1974) and need not be

repeated here except to note a persistent finding in these "fulfillment" studies has been that households reporting no intention or a zero purchase probability were later observed to account for a very substantial share, frequently the majority, of subsequent purchases (McNeil 1974, p. 8 and Juster 1974, pp. 12 and 14). In the context of Morrison's model, such systematic discrepancies could be interpreted as indications of instability (ρ) or bias (b), or some combination of the two.

Efforts to link intentions to purchase data for frequently purchased branded goods have been rare and generally too limited in scope to produce much insight into the sources of systematic errors. An exception is the work of Bird and Ehrenberg (1966) who conducted a detailed analysis of the relationship between intentions and usage using an extensive body of survey data collected in the United Kingdom covering more than 100 brands from 20 product categories. The intentions variable investigated was the percentage of respondents who selected a particular brand from a designated list when asked a straightforward question about which brand they were "likely to buy in the future". Current usage was similarly defined as the percentage of respondents who claimed to have used the brand within some specified time period. The bulk of their analysis focused on the concurrent relationship between intention and usage levels across brands within the same product category.

The finding reported by Bird and Ehrenberg (1966, p. 32) of particular interest here was that for brands with stationary usage levels, the percentage of respondents expressing an intention to buy exceeded the percentage claiming to be users-- "for all brands and product fields, irrespective of usage definitions."

Bird and Ehrenberg went on to show that the likelihood of a respondent expressing an intention to buy a brand was strongly associated with the recency of the respondent's past usage of that brand. As a "typical example", they presented data which showed that the proportion of respondents intending to buy a brand was .95 for current users of the brand, .45 for those who were not current users but had used the brand within the past 6 months, .10 for those

who had last used the brand more than 6 months ago, and .05 for those who had never used it. Bird and Ehrenberg note that purchase frequency will tend to be reflected in the reported recency of last use--i.e., recency of use will tend to be correlated with frequency of purchase. Thus the decline in the proportion expressing intentions as the time since last use increases implies some correlation between expressed intentions and purchase frequency. But it is clear from the above example that when the overall usage level is stationary over time, expressed intentions will overstate the actual frequency of purchase.

The fact that the likelihood of expressing an intention to purchase is associated with recency of past use suggests that this may be an example of how biases arise in judgements about uncertain events due to respondents' reliance on what Tversky and Kahneman (1974, p. 1130) call the "availability heuristic"-- "the tendency to assess the relative frequency of an event by the ease with which instances or occurrences can be brought to mind." Recent purchases are more likely to be recalled than previous ones and respondents may underestimate their interpurchase times and hence overestimate their frequency of purchasing. There is considerable evidence from other studies of the errors respondents make in reporting past events (as opposed to intentions) that for frequently occurring events, there is a tendency for the total number of events to be overreported (Sudman and Bradburn 1974, Chapter 3). This phenomenon is a result of "telescoping error" whereby an event is remembered as having taken place more recently than it actually occurred. Faced with the task of answering an intention question asked with reference to a particular brand of some frequently purchased product, a respondent may seek to make a judgement about the likelihood of purchasing the brand in the future based on his/her recall of past purchases. If such occurs, then we would expect that the functioning of the normal memory process would involve telescoping and thereby produce a response which overstates his/her true purchase frequency.

It is more difficult to conjecture about the judgemental heuristics underlying responses to intention questions for durable goods. The major known bias noted above, the tendency for those with no intention or a zero-purchase probability to make a substantial number of purchases, suggests that the bias is of the opposite sign and fundamentally different from that which appears to persist for packaged goods. The extended "post-mortem" reviews by McNeil (1974) and Juster (1974) of the Bureau of Census' experience with the Survey of Consumer Buying Expectations give the impression that reasons for this well-documented downward bias remain something of a mystery.

Dirkson and Wilkie (1978, p. 86) take a different view. They contend that "A survey of the literature...reveals several sensible explanations for the high incidence of 'unplanned' purchases and unfulfilled planned purchases" and go on to argue that "few durable purchases can be viewed as unplanned when contingencies, priorities, and situational constraints are considered."

In terms of the parameters of Morrison's model, it could be that the respondents with initially low or non-existent intentions are the group most likely to experience a positive change in true intention as a result of unforeseen events such as the unexpected breakdown of durables. Failure to consider contingencies under which one would replace an existing durable would lead to an underestimate of a purchase probability and could be an example of an "imaginability" bias (Tversky and Kahneman 1974, p. 1127). Neter (1970, pp. 16-17) reviews some limited evidence which suggests that recall of the purchase dates of a major appliance is subject to telescoping error. If so, respondents would tend to underestimate the age of an appliance and the probability of having to replace it. Clearly much remains to be learned here and a better understanding of the judgement heuristics underlying response biases could be useful in improving the design of measuring instruments.

2. Response Style

It has frequently been observed that in providing intention ratings for packaged goods stimuli, substantial numbers of respondents give identical or very similar response to the entire set of alternatives they are asked to rate. This tendency has generally been interpreted as a manifestation of some type of response style bias (or tendency to select a particular scale response category a disproportionate amount of the time, independent of the scale content or referent) such as acquiescence or yeasaying/naysaying (Wells 1961).

The problem is particularly troubling when intention scales are used as the criterion measure in concept tests. Several studies have found that approximately 30 per cent of respondents are "non-discriminators" with respect to their intention ratings of alternative stimuli (Clancy and Garsen 1970, Gold and Salkind 1974, Ptacek and Ross 1979). Research on psychological testing has suggested that response style tendencies are most likely to be operative when subjects are ambiguous or uncertain about how to respond (Cronbach 1950, Jackson 1967). Hence, intentions ratings obtained in concept tests may be especially prone to response style bias when respondents have only a limited exposure to the concept stimuli and/or when the concepts are highly similar.

The amount and source of variability of intentions ratings may be illustrated here by examining the results of some further analysis of the data from Davis' study described above in the discussion of reliability. In particular, we carried out an analysis of variance using the following mixed model for repeated measures:

$$X_{ict} = \mu + \alpha_i + C_c + T_t + (\alpha C)_{ic} + (\alpha T)_{it} + (CT)_{ct} + \epsilon_{ict} ,$$

where:

X_{ict} = respondent i's rating of concept c on occasion t
on the seven-point intention scale,

μ = grand mean,

- α_i = effect of respondent i , $i = 1, 2, \dots, 63$,
 C_c = effect of concept c , $c = 1, 2, \dots, 20$,
 T_t = effect of occasion t , $t = 1, 2$,
 $(\alpha C)_{ic}$ = effect of interaction of respondent i with concept c ,
 $(\alpha T)_{it}$ = effect of interaction of respondent i with occasion t ,
 $(CT)_{ct}$ = effect of interaction of concept c with occasion t ,
 ϵ_{ict} = residual.

The effects of concepts and occasions are taken as fixed while respondents are assumed to be random. The ANOVA summary is presented in Table 2. Note our use of the ANOVA results is confined to obtaining estimates of the contributions made by the different effects posited in the above model to the total variance of the observed intention ratings. The assumptions of normality and homoscedasticity of the error component required for F tests of hypothesis to be strictly valid and for estimates of the sampling errors of parameters to have optimum statistical properties will not in general hold for beta binomial distributed intentions ratings and are not satisfied for the particular set of data under consideration here. We are interested in the above ANOVA model as a means of summarizing the sample data rather than as a basis for making statistical inferences and tests. The variance components shown in Table 2 were obtained by setting the mean squares for the effects equal to their corresponding expected values and solving the resulting equations (Cronbach et al. 1972, Chapt. 2).

INSERT TABLE 2 HERE

We interpret the variance associated with the main effect of respondents as being largely attributable to individual differences in response styles,

although any differences among respondents with respect to some general "innovativeness" trait would also be reflected in this component. In any case, it is noteworthy in Table 2 that the variance component attributable to respondent differences across concepts is one and a half times greater than that due to mean differences among this highly diverse set of concepts. In a normal concept test, conducted on a single product category, one would expect that the set of stimuli investigated would be more homogeneous than the broad range of alternatives listed in Table 1 and hence an even smaller amount of variation due to between concept mean differences would likely be observed.

The largest variance component in Table 2 is that which is attributed to the interactions between respondents and concepts and accounts for almost half (46 per cent) of the total variance in the observed intention ratings. This indicates a substantial amount of idiosyncratic discrimination among concepts by respondents. It is interesting to note that the index Haley and Case (1979) used to compare the discriminant ability of several widely used rating scales is one which essentially defines discrimination solely in terms of the magnitude of the main effects of between stimuli mean differences. The type of design they employed does not allow any effects due to the interaction of respondents with stimuli to be separated from the residual variance. As suggested by Gold and Salkind (1974, p. 22), more information about the appeal of alternatives to sub-segments of respondents might be obtained from concept or brand discrimination tests employing intention scales if each respondent rated each stimulus twice rather than once which appears to be the usual practice.

Are More Scales Points Better?

The expression (11) for the reliability of beta binomial intention scales presented earlier in this paper implies that the reliability of such a scale can be increased by expanding the number of scale points it contains. Unfortunately,

no relevant evidence bearing directly on this issue appears to be available presently. In his extensive recent review of the literature dealing with the "optimal" number of response alternatives for a scale, Cox (1980, p. 48) concludes that "the empirical studies provide additional, consistent evidence suggesting that increasing the number of alternatives may increase the reliability of a scale but that the potential is probably minor in comparison with other means." However, the bulk of the evidence Cox discusses pertains to composite rather than single item scales such as intention measures.

Cox draws attention to another issue raised some years ago by Cronbach (1950) who maintained that the reduction in random measurement error obtained by expanding the number of response alternatives would be accompanied by a large component of systematic error due to response style bias. Cronbach (1950, p. 22) suggested that "the argument that the finer scale gives more reliability is not a sound one, since this is what we would expect if all of the added reliable variance were response-set variance and had no relation to the beliefs about the attitude object in question." If the advantage with respect to reliability of scales with more response alternatives is offset by an increase in response style bias, then their discriminant and predictive ability would also be adversely affected.

Some support for Cronbach's (1950) warning that scales with more response alternatives may be more afflicted by response style bias than scales with fewer levels is found in Haley & Case's (1979) study. Their results (p. 29) indicate that a five point intention scale discriminated somewhat better among brands than did the version of the eleven point purchase probability scale they employed. This directional result appeared in four of the six product categories investigated but in two of the aforementioned four categories the differences were

slight. Furthermore, in contrast to the original Juster scale and Wells's (1961) ten point "readiness to buy" scale, Haley and Case's purchase probability scale included descriptive labels for only four of the eleven response alternatives. As they suggest (p. 31) this feature of their eleven point scale may have diminished its discriminant ability.

Thus the relative merits of longer versus shorter intention scale remain an unresolved question. A well-designed study that investigates the interactions and tradeoffs between reliability and discriminant ability for scales of different lengths is needed to untangle the issues.

In general then, the evidence available from investigations of the discriminant quality of intentions ratings and from direct comparisons of intention and purchasing levels both indicate the presence of a sizable component of systematic error. While the sources of such biases are not well understood, the task with which respondents are confronted in reporting their intentions for a generic type of durable good would appear to be quite different than when asked to express their intentions with respect to a specific brand of some packaged good if only because of inherent differences in the length and regularity of interpurchase times for these two types of purchases. Thus it is of considerable interest to explore the application of Morrison's model over a broad range of purchase decisions. In the next section we present some further results obtained by fitting the model to data for both durable and packaged goods.

APPLICATIONS AND TESTS OF THE INTENTION-PURCHASE MODEL

Data

The essential data needed to estimate the parameters of Morrison's model are the marginal distribution of purchase intentions measured at one point in time and information about the incidence of subsequent purchasing by the same respondents tabulated according to the level of their previously expressed

intentions. A careful search of the published marketing research literature uncovered a few data matrices of the aforementioned type which were reported in sufficient detail to permit the re-analysis required to estimate and test Morrison's model. Table 3 lists the sources of the data used here and summarizes the pertinent features of each data set and the manner in which it was collected.

 INSERT TABLE 3 HERE

Estimation and Testing

The frequency distribution of intention scale responses is used to obtain maximum likelihood estimates of the beta binomial model parameters. Observed purchase frequencies for respondents with various stated intention values are used for estimating the parameters of the linear model relating purchase probability and stated intention. Finally, a likelihood ratio test is used to compare the goodness-of-fit of alternative models. A description of each of these estimation and testing procedures follows.

1. Estimation of the Beta Binomial Model

If the random variable x ($=0,1,\dots,n$) is distributed beta-binomial with parameters α and β , then the probability of observing $I_x = x/n$ is:

$$P(I_x) = \binom{n}{x} \frac{B(\alpha+x, \beta+n-x)}{B(\alpha, \beta)}, \quad (17)$$

where $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ and $\Gamma(\cdot)$ denotes the gamma function. Let N_x represent the number of households with a response of x on a stated intention scale ranging from 0 to n . Then, the likelihood function for observing frequencies N_x 's is given by:

$$L(N_x's; \alpha, \beta) = \prod_{x=0}^n P(I_x)^{N_x}. \quad (18)$$

Constant terms ignored, the logarithm of the likelihood function simplifies to:

$$L(N_x's; \alpha, \beta) = \sum_{x=0}^n N_x \left\{ \sum_{r=0}^{x-1} \log(\alpha+r) + \sum_{r=0}^{n-x-1} \log(\beta+r) - \sum_{r=0}^{n-1} \log(\alpha+\beta+r) \right\} \quad (19)$$

The maximum likelihood estimates of α and β can be obtained by maximizing the log-likelihood function in equation (19) through numerical optimization (Kalwani 1980). Standard errors of the maximum likelihood estimators are obtained from the inverse of the information matrix using the Cramer-Rao lower bound for the asymptotic variance of estimators (Rao 1965, pp. 289-302).

2. Estimation of the Linear Model

As noted above, Morrison (1979) proposed a linear model for the relationship between purchase probability and stated intention. That is:

$$P_x = A + BI_x \quad , \quad (20)$$

where P_x is the purchase probability of a respondent with stated intention I_x . Equation (18) indicates that the intercept and slope terms are given by:

$$A = \frac{\rho\alpha}{\alpha+\beta} + \frac{(1-\rho)\alpha}{\alpha+\beta+n} - b \quad , \quad (21)$$

$$B = \frac{(1-\rho)n}{\alpha+\beta+n} \quad , \quad (22)$$

where ρ is the probability that there is a change in an individual's true intention and b is the systematic bias.

Let N_x represent the number of respondents with purchase intention I_x who buy the product and $m_x = N_x - n_x$ the number who do not buy the product. Then, the likelihood function for observing the purchase frequencies n_x 's and m_x 's is given by:

$$\ell(n_X's, m_X's; A, B) = \prod_{x=0}^n P_X^{n_X} (1-P_X)^{m_X}, \quad (23)$$

where $P_X = A + BI_X$. Taking logarithms, we obtain the log-likelihood function:

$$L(n_X's, m_X's; A, B) = \sum_{x=0}^n \left\{ n_X \log P_X + m_X \log (1-P_X) \right\}. \quad (24)$$

We find the maximum likelihood estimates of A and B by maximizing the log-likelihood function in equation (24) through numerical optimization. Standard errors of maximum likelihood estimators are obtained, as before, from the inverse of the information matrix using the Cramer-Rao lower bound for the asymptotic variance of estimates (Rao 1965, pp. 289-302).

Equation (22) is used to obtain the maximum likelihood estimate of ρ from the maximum likelihood estimates of α , β and B, that is:

$$\hat{\rho} = 1 - \hat{B} \frac{(\hat{\alpha} + \hat{\beta} + n)}{n}. \quad (25)$$

If b is estimated from equation (16), then our model suggests that the estimated value of A should be given by equation (21).

Tests of Alternate Models

We use the likelihood ratio test to compare the goodness-of-fit of alternate models. First, we compare the linear model:

$$P_X = A + BI_X, \quad (26)$$

with the null hypothesis $P_X = A$ with B set equal to zero. For this purpose, we form the ratio of the maximum value of the likelihood function of the unconstrained model under the null hypothesis (say, model 0) to that for the unconstrained model (say, model 1). That is:

$$\lambda_{01} = \frac{\ell^*(0)}{\ell^*(1)},$$

where the asterisk denotes that it is the maximum value. For large samples, Wilks (1962, Chapter 13) has shown that $-2\log_e \lambda$ is chi-square distributed with 1 degree of freedom representing the difference in the number of the parameters of the unconstrained and constrained models.

We also compare the linear model with the piecewise linear model. The mathematical form of the latter model is:

$$P_x = A + B_1 I_x, \quad \text{if } I_x \leq I_x^*, \quad (27A)$$

$$P_x = A + B_1 I_x^* + B_2 (I_x - I_x^*), \quad \text{if } I_x > I_x^*. \quad (27B)$$

Figure 4 provides a graph of the piecewise linear model. The linear model given in equation (26) is a constrained version of the piecewise linear model in equation (27). The chi-square statistic from the likelihood ratio test in this case has $2(=4-2)$ degrees of freedom.

=====

INSERT FIGURE 4 HERE

=====

In this section, we present empirical results from analyses of stated intentions and purchase data for durable and non-durable goods. For each product we report parameter estimates and their standard errors for the beta binomial model describing purchase intentions (i.e., $\hat{\alpha}$, $\hat{\beta}$, $SE(\hat{\alpha})$ and $SE(\hat{\beta})$) and the linear model linking stated intentions to purchase probabilities (i.e., $\hat{A}$, $\hat{B}$, $SE(\hat{A})$, and $SE(\hat{B})$). As indicated previously, the maximum likelihood estimate of the parameter ρ representing the probability of change in a given individual's true intention is obtained from $\hat{\alpha}$, $\hat{\beta}$ and $\hat{B}$ using equation (22).

We employ four different performance measures to evaluate the models--linear and piecewise--linking stated intentions to purchase probabilities. The first measure is a chi-square statistic obtained from the likelihood ratio test as described in the previous section. Non-significant values of the statistic indicate that the null hypothesis cannot be rejected in favor of the alternate hypothesis. We perform two likelihood ratio tests for each durable and packaged good data set. In one case we treat the linear model as the alternate hypothesis with $P_x = A$ as the null model and in the other case we posit the piecewise linear model as the alternate hypothesis with the linear model as the null hypothesis.

The second performance measure is denoted by R^2 and is defined as in regression analysis as the ratio of "explained variance" to "total variance". It is given by:

$$R^2 = 1 - \frac{\sum_{x=0}^n N_x (p_x - \hat{p}_x)^2}{\sum_{x=0}^n N_x (p_x - \bar{p}_x)^2}, \quad (28)$$

where N_x = the number of respondents with stated intention $I_x = x/n$,

n_x = the number of respondents with stated intention who purchase the product,

$$p_x = n_x / N_x,$$

$$\bar{p}_x = \left(\sum_{x=0}^n n_x \right) / \left(\sum_{x=0}^n N_x \right),$$

$\hat{p}_x$ = the predicted purchase probability given that the stated intention is I_x .

The numerator of the second term on the right-hand side of equation (28) is, of course, the unexplained variance and the denominator is the total variance. As in regression, R^2 is a measure of the goodness-of-fit of the hypothesized model--linear or piecewise--to the observed purchase data.

We also report the standard error of estimate, S , in each case. Recall that S^2 is defined as:

$$S^2 = \frac{\sum_{x=0}^n N_x (p_x - \hat{p}_x)^2}{N - k - 1}, \quad (29)$$

where $N = \sum_{x=0}^n N_x$, k is the number of model parameters, and the other terms are defined as in equation (28). The standard error of estimate provides a measure of the error associated with predicting individual purchase probabilities and may be used to construct the confidence interval around the predicted values.

The final performance measure, we report, is denoted by R_M^2 and is defined as (see McFadden 1970, p. 121) :

$$R_M^2 = 1 - \frac{L^*(A)}{L^*(N)}, \quad (30)$$

where $L^*(A)$ and $L^*(N)$ represent the maximum values of the log-likelihood functions for the alternate and null models respectively. The measure R_M^2 takes the value one when $L^*(A)$ is zero which happens when the likelihood function is a product of unit probabilities. In other words, R_M^2 takes the value one when with the alternate model, we are able to predict outcomes (i.e., purchase vs. non-purchase)

with probability one. If the purpose of models linking prior intentions to purchase probabilities is to compare the overall sample response to alternate stimuli, then R_M^2 is a less appropriate performance measure than R^2 . We report R_M^2 here for completeness. Below we present the results from data on durable and branded packaged goods.

Durable Goods Results

In the case of durable goods, we report results for two sets of data. The first data set is from Juster (1966) and was used by Morrison (1979) to illustrate the application of his model. The source of the second data set is Pickering and Isherwood (1974). Table 4 displays the parameter estimates along with the four performance measures for the Juster data. Estimates of the beta binomial model parameters α and β were obtained from the stated intention frequency data, namely the N_x 's.

=====
INSERT TABLE 4 HERE
=====

Table 5 reports the actual and predicted purchase intention frequencies (rounded to the nearest integer) for the Juster data. Predicted frequencies were obtained from $\hat{\alpha}$ and $\hat{\beta}$. Examination of the results in Table 5 reveals that the beta binomial model provides a very good fit to the stated intention data in all four cases. In three of the four cases, the χ^2 statistics are less than the critical value of 26.125 (with 8 degrees of freedom and 1 percent significance level).

=====
INSERT TABLE 5 HERE
=====

In addition to the estimates of α and β , Table 4 also displays the findings from fitting the linear model to the Juster data.

Examination of the χ_1^2 values reveals that they are all greater than 6.635 which means that the probability of obtaining each of these χ_1^2 values is less than 0.01. Therefore, we reject the null hypothesis that predicted purchase probabilities are independent of stated intentions. We conclude that the linear model, $P_x = A + BI_x$, provides a superior fit to the data than the null model, $P_x = A$.

The values of R^2 provide an indication of the goodness-of-fit of the linear model. Figures 5 through 8 provide a graphical display of the quality of the fit. It is worthwhile noting that for automobiles, 12 months intention data (Figure 6) with $R^2 = 0.972$ and standard error, $S = 0.027$ fit six-month purchases better than do six-month intentions data (Figure 5) with $R^2 = 0.772$ and standard error, $S = 0.069$. For household appliances, six-month intentions data were not obtained but twelve-month intentions data (Figure 7) with $R^2 = 0.890$ and standard error, $S = 0.009$ fit six-month purchase data better than do twenty-four month intentions data (Figure 8) with $R^2 = 0.752$ and standard error, $S = 0.019$. As Morrison (1979, p. 72) notes, this raises the question of the optimum time horizon (say, x months) for collecting intentions data given that purchase predictions over a period of certain length (say, y months) are desired. The analysis of the stated intentions from Juster (1966) reported in Morrison (1979) clearly indicates that respondents are sensitive to the time horizon. For both automobiles and household appliances, there is a shift in the frequency distribution of stated intentions from a lesser towards a greater intent to purchase the product ($I_x=0$ to $I_x=1$) as the time horizon increases.

=====

INSERT FIGURES 5-8 HERE

=====

The standard errors of estimate are given in the second to last column of Table 4. As an illustration of their implication, consider the data set, Autos II. The predicted purchase probability for those respondents with $I_x = 2/10$ is

$$\hat{P}_x = \hat{A} + \hat{B}I_x = 0.071 + 0.528(2/10) = 0.177.$$

Assuming $\hat{P}_x$ is normally distributed, the 95% confidence interval around $\hat{P}_x$ is given by $\hat{P}_x \pm 2S = 0.177 \pm 2(0.027)$, or, 0.123 to 0.231.

The estimates of ρ in Table 4 are obtained from equation (25). Note that ρ 's for automobiles are smaller than those for household appliances indicating that respondents' intentions to buy automobiles are more stable than in the case of appliances.

Finally, Table 4 also displays numerical values of bias b due to systematic errors for automobile and appliance data. It is estimated as follows:

$$b = \frac{\hat{\rho} \hat{\alpha}}{\hat{\alpha} + \hat{\beta}} + \frac{(1-\hat{\rho}) \hat{\alpha}}{\hat{\alpha} + \hat{\beta} + n} - \hat{A}.$$

Positive values of b indicate that the mean of expressed intentions exceeds the observed proportion of purchases. It is worth noting that the maximum likelihood estimates of A are stable; that is, the associated standard errors of estimate yield t -values of over 4.6. This finding indicates that, in case of durable goods, the assumption of constant bias across respondents is not unreasonable since to the extent that systematic errors due to response set tendencies or other sources vary across respondents we would observe large variation in A .

Table 6 displays the results for the Pickering and Isherwood data. Estimates of the beta binomial model parameters α and β are given in columns 3 and 4. For space reasons, we do not report the observed and predicted stated intention frequencies for the Pickering and Isherwood data. None of the computed χ^2 values exceeded the critical value of 26.125 (with 8 degrees

of freedom and 1 percent significant level) and the quality of fit was comparable to that of the Juster data.

All χ^2_1 statistics (except that for black and white television) in Table 6 are significant at the 1 per cent level. The sole exception of $\chi^2_1 = 3.313$ in the case of black and white television is significant at the 10 per cent level. Therefore, we reject the null hypothesis that $P_x = A$. Once again, we conclude that for durable goods the linear model provides a superior fit to the data than the null model, $P_x = A$. The R^2 values are reasonably high with the median $R^2 = 0.535$. Numerical values of standard errors of estimate and R_M^2 are comparable to those of the Juster data.

As in the case of the Juster data, we find that the estimates of A are stable (i.e., the t-values are statistically significant). Once again, this finding indicates that the assumption of constant bias is tenable in the case of durable goods.

INSERT TABLE 6 HERE

A major finding from the durable goods results reported in Tables 4 and 6 is that the linear model provides a significantly better fit than the null model ($P_x = A$) with purchase probabilities being independent of intentions. A natural question that arises is whether there are other models of a relatively simple form which would provide better fits to the data than does the linear model. One reasonable alternative non-linear model is the piecewise linear model. Its mathematical form is given in equation (27) and a graphical representation is shown in Figure 4. As before, we employ the likelihood ratio test to compare the goodness-of-fit of the linear and piecewise linear models.

Results of this comparison for the Juster data are displayed in Table 7. In all four cases the null hypothesis cannot be rejected at the 1 per cent levels. We conclude that the linear model provides a superior fit to the

Juster data than the piecewise linear model. The improvement in R^2 (and R_M^2) in going from the linear to the piecewise linear model is small. Similarly, the reduction in standard error of estimate is small.

=====

INSERT TABLE 7 HERE

=====

We also carried out a similar comparison of the linear and piecewise linear models for the Pickering and Isherwood durable goods data. The results are similar to those of the Juster data and are presented in Table 8. The χ^2 statistics obtained from the likelihood ratio tests are not significant at the 1 per cent level in any of the ten cases. The improvements in R^2 (and R_M^2) and the reductions in the standard error of estimate are small. From all these findings, we conclude that for durable goods, the linear model constitutes a reasonable hypothesis for relating purchase intentions to purchase probabilities.

=====

INSERT TABLE 8 HERE

=====

Packaged Goods Results

We report results for five packaged goods. The sources of the data investigated here are listed in Table 3 which also provides some descriptive information about the studies from which they were obtained. Table 9 displays the parameter estimates along with the four performance measures for the five packaged goods.

=====

INSERT TABLE 9 HERE

=====

The estimates of the beta binomial model parameters α and β , as in the case of durable goods, were obtained from the stated intention frequency data,

N_x 's. Table 10 reports the actual and predicted intention frequencies (rounded to the nearest integer) for the five packaged goods. Predicted frequencies were obtained from $\hat{\alpha}$ and $\hat{\beta}$. Examination of the results in Table 10 reveals that the beta binomial model does not provide a fit of comparable quality to that obtained for the durable goods data. In three of the five cases, the χ^2 statistics are greater than the critical values at the 1 per cent significance level. In the case of Products B and C from Tauber (1975), however, the fit of the beta binomial model is reasonably good. In summary, the evidence for packaged goods presented here with regard to the beta binomial hypothesis is inconclusive.

=====

INSERT TABLE 10 HERE

=====

The χ^2 values from the likelihood ratio test comparing the fit of the linear model and the null model, $P_x = A$, in Table 9 all exceed 6.635 and, therefore, the null hypothesis that the predicted purchase probabilities are independent of stated intentions is rejected at the 1 per cent level. The values of R^2 are reasonably high suggesting that the linear model provides a fairly good fit for the packaged goods data.

Table 11 displays parameter estimates along with the performance measures for the piecewise linear model. The chi-square statistic has 2 degrees of freedom and is obtained from the likelihood ratio test with the linear model as the null hypothesis and the piecewise linear model with a threshold effect (see Figure 3) forms a reasonable alternative hypothesis for obtaining purchase probabilities from stated intentions.

=====

INSERT TABLE 11 HERE

=====

The numerical values of the estimates of A , B_1 , I_X^* , and B_2 in the piecewise linear model are interesting. Note that the maximum likelihood estimate of the threshold, I_X^* , was obtained from the optimization algorithm. In other words, maximum likelihood estimates of all four parameters were estimated simultaneously (Kalwani 1980). The estimate of slope 1 (B_1) is considerably smaller than the estimate of slope 2 (B_2). In three of the five cases, B_1 is equal to zero. The estimate of the intercept is also close to zero, which indicates that in the case of packaged goods, the predicted purchase probability of those with stated intentions less than the threshold is close to zero. Perhaps then, for these respondents the branded packaged good is part of their evoked set but they are not seriously considering buying the brand. It bears noting that in four of the five cases (the exception is Product B from Tauber (1975)), only those with stated intentions of $(n-1)/n$ and n/n have a significant probability of purchasing the product. This finding lends support to the rule-of-thumb frequently employed in analyzing concept tests wherein only the responses for the "top two boxes" of 5-point intention scales are used as a basis for concept selection.

DISCUSSION

The major finding that emerges from the results presented in the preceding section is that the relationship between expressed intentions and subsequent purchase behavior appears to be different for generic durables as compared to branded packaged goods. Whereas intentions and purchasing were found to be linearly related for durable goods, the packaged goods data exhibited a threshold

effect and the slope of the intention-purchase relationship was not constant. For the entire set of durable goods investigated, the linear model provided a superior fit to the data than either the null-model that purchase probabilities are independent of purchase intentions or the piecewise linear model. In contrast, for packaged goods the linear model gave an inferior fit to that obtained with the piecewise linear model. Moreover, the estimates of the intercept and slope 1 in the piecewise model were found to be close to zero indicating that respondents with intentions below the threshold level are unlikely to buy branded packaged goods. The t-statistics for the parameter estimates and several other performance measures all indicated that the linear model fitted the durable goods data quite well as did the piecewise linear model for the packaged good brands.

It should be mentioned that a threshold effect also appears in results reported by Haley and Case (1979, p. 30) and Laroche and Howard (1980, pp. 382 and 384) summarizing concurrent (as opposed to predictive) relationships between intentions and purchasing for packaged good brands. Laroche and Howard employed a five point intention scale while Haley and Cases's data show a threshold effect for both the five point intention scale and the eleven point purchase probability scale they included in their study. The existence of a threshold effect is consistent with the "top box" rule sometimes used to score intention ratings whereby only respondents who check one of the two highest response categories are counted in computing a criterion measure in concept tests and related types of studies (Gold and Salkind 1974). As has been previously suggest by Gold and Salkind (1974), a likely explanation for the threshold effect is yeasaying response style bias. Another not necessarily unrelated interpretation suggested by Tversky and Kahnman's (1974) concept of how bias

arises in judgments due to reliance on an "availability" heuristic would be that recency and/or frequency of purchase plus telescoping memory error may be the source of bias in intention ratings for packaged goods.

In line with our earlier discussion of bias, the results obtained with the linear model suggest that the systemic error in individual intentions for generic durables may arise from different sources than for brands of frequently purchased products. Referring to the results for the Pickering and Isherwood data summarized in Table 6, we see the estimates of the bias parameter, $\hat{b}$, varied in both sign and magnitude: $\hat{b}$ was positive for 7 of the 10 durables and ranged in absolute value from .009 to .065. These results plus the variability in both the beta binomial parameters suggest more discrimination among the different durables than would be expected if a large component of response style bias was consistently present in the intention ratings for this entire set of durables. Recall that the same sample of respondents rated all 10 durables. On the other hand, we observe from Table 6 that the change parameter, ρ , was generally quite high, ranging from .298 for new automobiles to .881 for black and white television. All but two of the estimates of ρ exceeded .5 implying that for most of these durables, half or more of the respondents underwent a change in true intentions at some point in the 12 month period covered by the study.

The beta binomial model was consistently found to fit the marginal distribution of intentions ratings quite well for a broad spectrum of generic durable goods but this was not the case for packaged goods where the scope of our investigation was restricted to a handful of brands. For only two of the five sets of package goods data examined could the null hypothesis of the beta binomial model be sustained. However, in the two instances where the fit was clearly poor (Smith et al and Twyman), the intentions data analyzed consisted of ratings for more than one product that had been pooled. The poor fits obtained for the beta binomial model may be the result of aggregating hetero-

geneous underlying distributions of the individual products. Thus we must regard the evidence reported here bearing on the quality of the fit of packaged goods data to the beta binomial model as inconclusive.

It should be mentioned that our conclusions pertaining to the comparative performance of alternative forms of the intention-purchase probability relationship (independence vs. linear vs. piecewise linear) do not in any way rely on the assumption that stated intentions follow a beta binomial distribution. As discussed in Morrison (1979, pp. 67-68), the beta binomial assumption along with the modifications allowing for instability and bias simply represent behavioral assumptions which are consistent with the hypothesis of a linear intention-purchase relationship. Therefore, the poor fit of the linear model in the case of branded packaged goods may not be attributed to the failure of the stated intentions data to fit the beta binomial model.

In his original paper, Morrison (1979, p. 72) suggested that further empirical applications of this model could lead to the discovery of systematic differences in the properties of intentions data that would relate to the types of purchase stimuli for which intentions were being measured. Our findings indicate that such a distinction does exist between generic or broadly defined classes of durable goods purchased infrequently and specific brands of packaged goods from product categories where purchases occur frequently and regularly. Needless to say, such a comparison represents only a modest beginning in developing an understanding of the properties of intention scales that is needed to address the issues pertaining to their usefulness that arise in practice when heavy reliance is placed on an intention scale as the key response or criterion measure. Our search of the published literature turned up no detailed evidence relating to the reliability and/or predictive validity of intentions measures for specific brands of durables. Similarly, other critical distinctions such as that between new versus established brands and concepts versus

existing products remain to be investigated as does the issue of parameter stationarity over time to which Morrison (1979, p. 73) drew attention. Doubtless, experience and data relevant to some or all of these questions has accumulated in the marketing research community and this paper will have served a valuable purpose if it stimulates further applications of Morrison's model and disclosure of the results.

SUMMARY AND CONCLUSIONS

In this paper we have shown how Morrison's model can serve as a valuable framework for organizing existing evidence bearing on the key psychometric properties of intention scales: unreliability, bias, and instability. Each of these qualities represents a potential threat to the predictive validity of intentions measures and is explicitly represented in the model. Our review of previously published findings plus the results of the analysis reported here may be summarized as follows:

1. Reliability

Using Morrison's assumptions that stated intentions follow a beta binomial distribution, we showed that reliability is expected to vary according to the heterogeneity of true intentions among respondents and the number of scale points included in the response scale. Findings presented from a small scale study of a seven point instrument indicated that the test-retest correlations for intention ratings of concept stimuli were consistent with the levels of reliability expected under the beta binomial assumptions. There is a conspicuous absence of other studies of the reliability of intention scales in the marketing research literature.

2. Bias, Instability, and Predictive Validity

Evidence of the predictive validity of intention ratings was obtained from re-analyses of several sets of previously published data but different forms of predictive relationships were found for generic durable goods and branded packaged goods. There is reason to believe that the sources of systematic error present in intention ratings are different for these two types of purchases.

a. Generic Durable-Good Purchases

Consistent with Morrison's model, stated intentions were found to fit beta binomial distributions and be linearly related to purchase probability for a broad spectrum of durable goods purchases. The estimated value of the change parameter was generally quite high and implied that a very substantial fraction of respondents experienced a change in true intention over a twelve month period. The magnitude of bias was generally positive but varied in magnitude over a limited range. The assumption of constant bias across different levels of intention also appeared tenable.

b. Branded Packaged Good Purchases

A threshold effect appears to characterize the intention-

purchase relationship here and a piecewise linear model was found to provide a good fit to the limited number of data sets investigated. The probability of purchase appears to be quite small for intentions below the threshold level and is consistent with the "top box" scoring rule frequently applied to intention ratings for packaged goods brands and concepts. Response style and judgment heuristics were suggested as likely explanations of the threshold effect. Results obtained pertaining to the fit of beta binomial model were inconclusive.

3. Scale Design

There are reasons for hypothesizing that increasing the number of scale points or response alternatives included in an intention instrument may increase both reliability and response bias. There is insufficient empirical evidence to evaluate these effects and thus resolution of the issue of optimal scale length awaits further study.

These findings provide some reinforcement for Morrison's (1979, p. 73) belief that "empirical regularities" would emerge from further studies of intentions data. Hopefully the methods and results discussed here will encourage further applications of Morrison's model to expand the scope of present knowledge about this widely used measure.

APPENDIX A

In this appendix we obtain an expression for the two variance components of a beta binomial distributed intention scale, namely, the variance of the conditional mean and the expected value of the conditional variance. The first variance component is given by:

$$\text{Var}[I_t] = \int_0^1 I_t^2 f(I_t) dI_t - \{E[I_t]\}^2, \quad (A1)$$

where $f(I_t)$ is the beta density function with parameters α and β . Upon integration, equation (A1) leads to:

$$\begin{aligned} \text{Var}[I_t] &= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \left(\frac{\alpha}{\alpha+\beta}\right)^2 \\ &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \end{aligned} \quad (A2)$$

The second variance term representing the expectation of the conditional variance is given by:

$$E[\text{Var}[\varepsilon_t | I_t]] = \int_0^1 \frac{I_t(1-I_t)}{n} f(I_t) dI_t. \quad (A3)$$

Upon integration, equation (A3) becomes:

$$E[\text{Var}[\varepsilon_t | I_t]] = \alpha\beta/n(\alpha+\beta)(\alpha+\beta+1). \quad (A4)$$

From equations (A2) and (A4), the total variance of I_x is given by:

$$\begin{aligned} \text{Var}[I_x] &= \text{Var}[I_t] + E[\text{var}[\varepsilon_t | I_t]] \\ &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} + \frac{\alpha\beta}{n(\alpha+\beta)(\alpha+\beta+1)} \\ &= \frac{\alpha\beta(\alpha+\beta+n)}{n(\alpha+\beta)^2(\alpha+\beta+1)}. \end{aligned} \quad (A5)$$

APPENDIX B

The purpose of this appendix is to derive an expression for the linear relationship between the expected value of true intention and the stated intention. The random variable, I_x , representing the stated intention is distributed beta-binomial such that:

$$P(I_x) = \int_0^1 \binom{n}{x} I_t^x (1-I_t)^{n-x} f(I_t) dI_t$$

$$= \binom{n}{x} \frac{B(\alpha+x; \beta+n-x)}{B(\alpha, \beta)}, \quad (B1)$$

where $B(\cdot)$ denotes the beta function. The conditional distribution of I_t for given I_x is:

$$P(I_t | I_x) = \frac{\binom{n}{x} I_t^x (1-I_t)^{n-x} f(I_t)}{P(I_x)}. \quad (B2)$$

From (B2) we obtain:

$$E[I_t | I_x] = \int_0^1 I_t P(I_t | I_x) dI_t$$

$$= \frac{1}{P(I_x)} \int_0^1 I_t \binom{n}{x} I_t^x (1-I_t)^{n-x} f(I_t) dI_t$$

$$= \frac{B(\alpha+x+1; \beta+n-x)}{B(\alpha+x, \beta+n-x)}.$$

Upon simplification we obtain:

$$E[I_t | I_x] = \frac{\alpha+x}{(\alpha+\beta+n)}$$

$$= \left(\frac{\alpha}{\alpha+\beta+n} \right) + \left(\frac{n}{\alpha+\beta+n} \right) I_x, \quad (B3)$$

since $x = nI_x$.

FOOTNOTES

1. Intention is also a key construct in models of the relationship between attitudes and behavior. Recent discussions of this work in social psychology and marketing may be found in Bagozzi (1981) and Warshaw (1980).
2. Keats and Lord (1962, p. 61) show that the reliability of beta binomial distributed scales (11) is equivalent to the Kuder Richardson Formula 21 (Lord and Novick 1968, pp. 91-92 and 523-524) used in psychometrics to assess the reliability of composite scores.
3. Another source of systematic error in the predictive ability of measured intentions that deserves mention here is that which arises when purchases are the outcome of a family decision process, which is not reflected in the intention measures obtained from an individual family member. Davis and Ragsdale (1979) and Granbois and Summers (1975) have reported low to moderate levels of correlation between purchase intention measures obtained from husbands and wives within the same households.

FIGURE 1
MODEL STRUCTURE

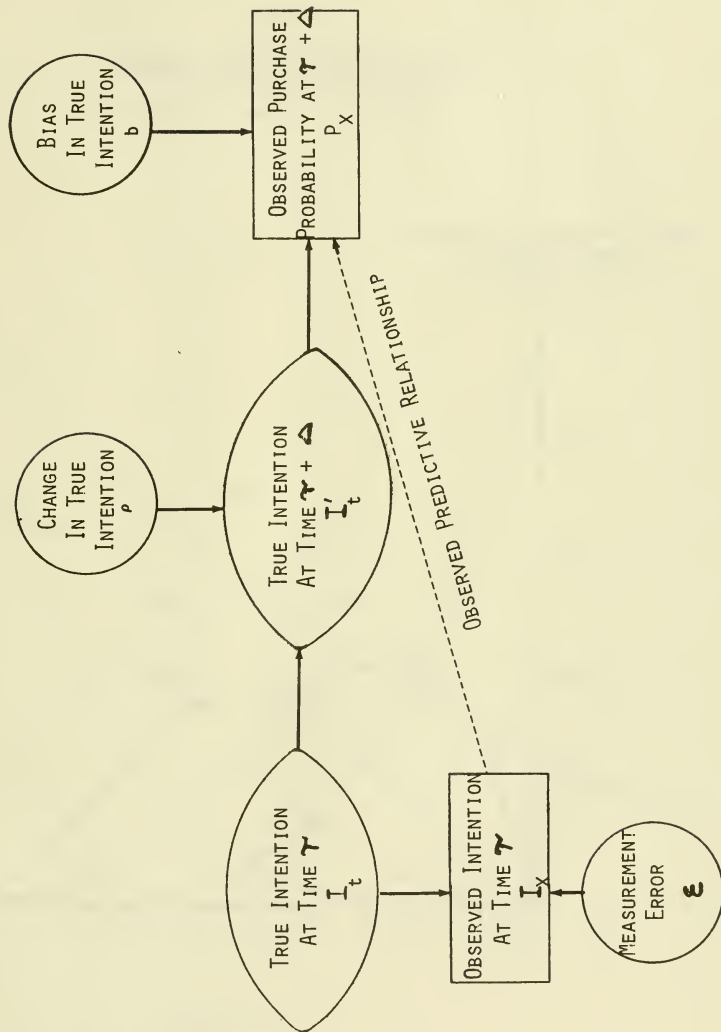


FIGURE 2

TRACE LINES FOR DIFFERENT RESPONSE
CATEGORIES OF A 5 POINT INTENTION SCALE

Binomial Probability of Endorsing
Response Category x
of a $n+1$ point scale
 $P(x; I_t, n=4)$

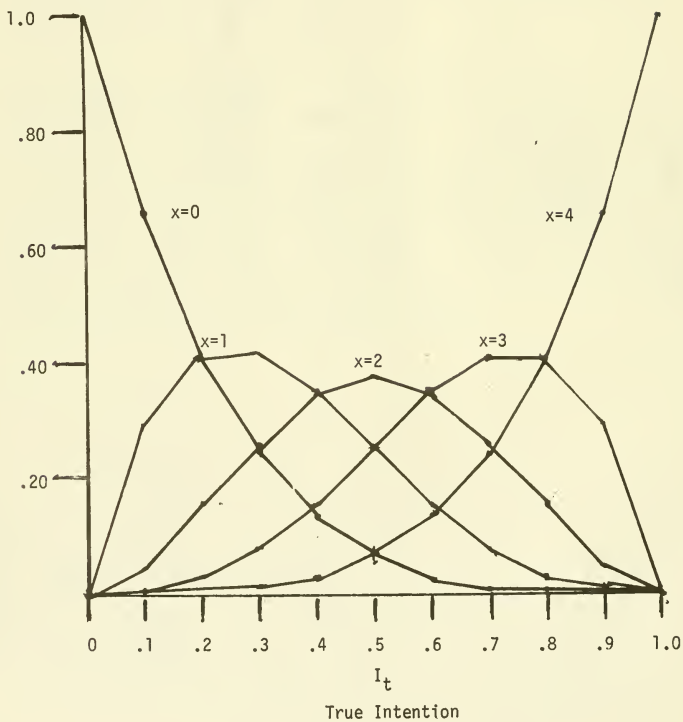


Figure 3

EXPECTED RELIABILITY AND OBSERVED TEST-RETEST
CORRELATION AS A FUNCTION OF $\alpha+\beta$ FOR A
SEVEN POINT INTENTION SCALE

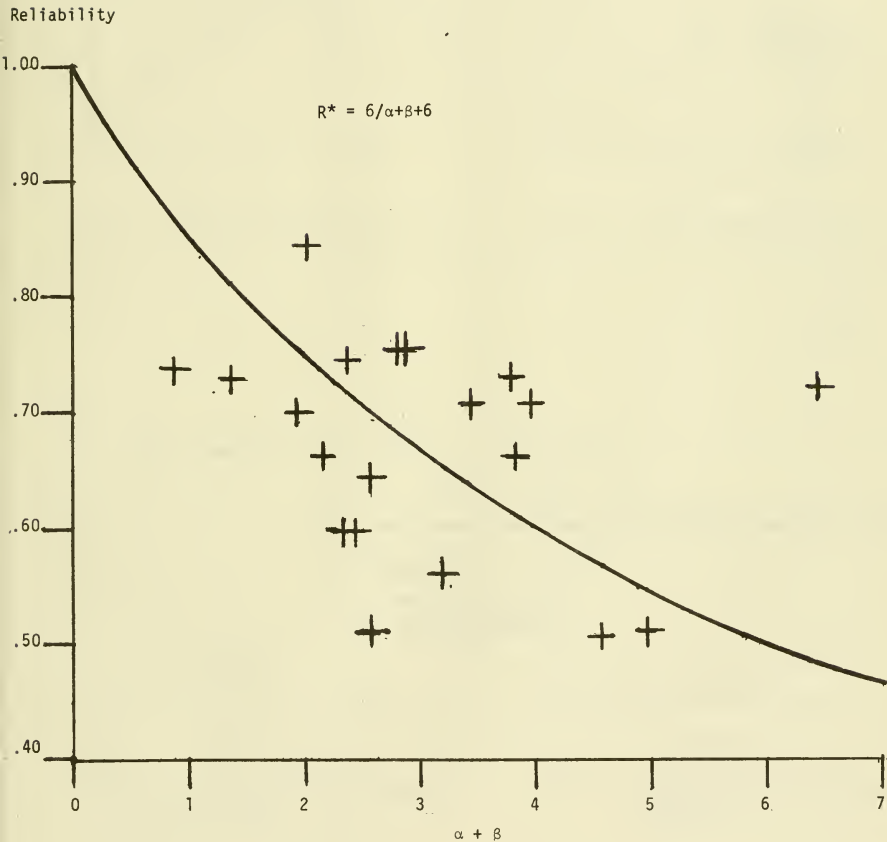
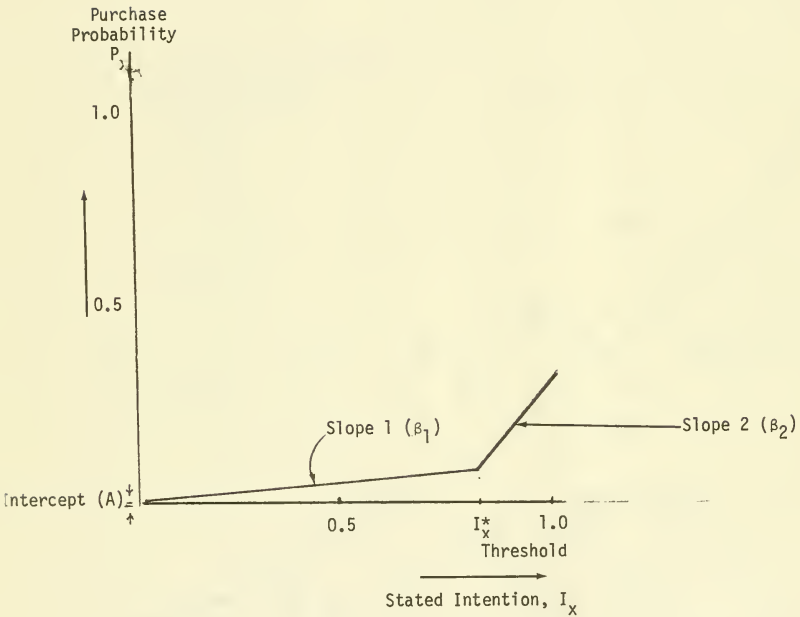


FIGURE 4

PIECEWISE LINEAR MODEL WITH A THRESHOLD AT STATED INTENTION, I_x^*



PREDICTED VS. ACTUAL PURCHASE PROBABILITIES FOR JUSTER DATA

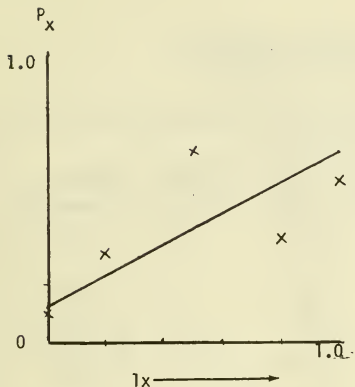


Figure 5: 6 month intentions vs.
6 month purchases of automobiles

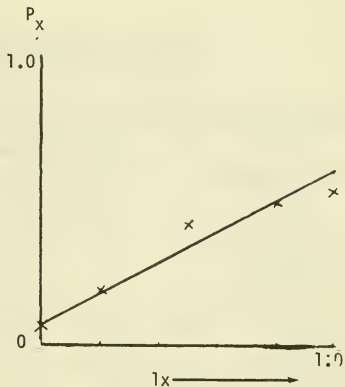


Figure 6: 12 month intentions vs.
6 month purchases of automobiles

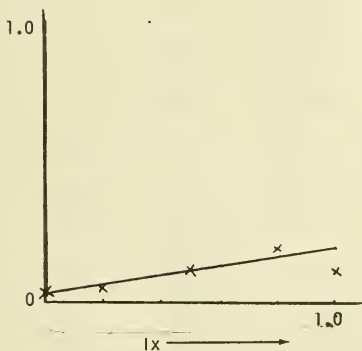


Figure 7: 12 month intentions vs.
6 month purchases of household
appliances

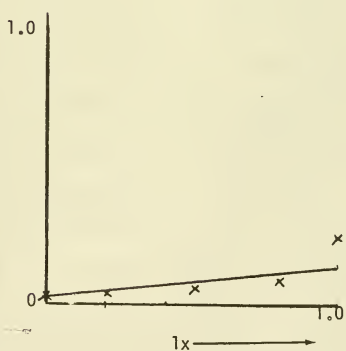


Figure 8: 24 month intentions vs.
6 month purchases of household
appliances

TABLE 1

BETA BINOMIAL PARAMETER ESTIMATES, EXPECTED RELIABILITY
AND OBSERVED TEST-RETEST CORRELATION FOR RATING OF
TWENTY CONCEPTS OF A SEVEN POINT INTENTION SCALE

Concept	Beta Binomial Parameter Estimates for Ratings Obtained on First Test*		Goodness-of-Fit Test for Beta Binomial Distribution† χ^2 (df=4)	Expected Reliability ($n/\hat{\alpha}+\hat{\beta}+n$)	Observed Test-Retest Correlation
	$\hat{\alpha}$	$\hat{\beta}$			
Energy-Saving Window Air Conditioner	.543(.097)	.378(.069)	5.857	.867	.737
Auto-Train Service	.754(.126)	.626(.106)	6.815	.813	.729
Do-it-Yourself Plumbing Fixtures	.895(.141)	1.958(.173)	5.320	.754	.700
Large Screen Color Television via Closed Circuit TV	1.115(.177)	.942(.149)	5.155	.745	.840
Investment Counseling	1.135()	1.030()	13.941	.735	.663
Easy-to-Install Sauna	1.262(.195)	1.063(.167)	6.140	.721	.596
Central Home Vacuum Cleaner	1.121(.174)	1.229(.192)	6.078	.719	.743
Natural Food Candy Bar	1.135(.172)	1.317(.207)	6.002	.710	.595
Economy Airfare (Limited Services)	.720(.116)	1.860(.329)	8.086	.699	.642
Disposable Fashion Raincoat	1.988(.362)	.607(.106)	4.884	.698	.507
Electric Powered Automobile	1.045(.159)	1.743(.277)	4.980	.683	.758
Automatic Bank Tellers in Supermarkets	.975(.151)	1.908(.308)	9.546	.675	.754
Combination Clothes Washer and Dryer	1.439(.211)	1.778(.266)	11.916	.651	.559
Art Rental Service	2.220(.341)	1.197(.180)	2.602	.637	.705
Home Burglar Alarm System	1.620(.231)	2.180(.319)	3.955	.612	.732
Pop-Top Cans for Soups, Vegetables, etc.	2.016	1.820	13.638	.610	.663
Hands-Free Telephone	1.762(.249)	2.220(.321)	14.482	.601	.709
Home-Shopping Service via Telephone Connected TV	2.034(.292)	2.568(.359)	0.917	.566	.506
Disposable Cheese-Board & Dinner Cheese Assortment	2.741(.376)	2.247(.309)	5.727	.546	.510
Video-Tape Unit for Home Education	2.623(.347)	3.871(.521)	10.590	.480	.722

*The figures in parentheses are the estimated standard errors of $\hat{\alpha}$ and $\hat{\beta}$.

†The critical values of the chi-square statistic (df=4) are:

$\chi^2_{.10} = 7.779$; $\chi^2_{.05} = 9.488$; and $\chi^2_{.01} = 13.277$.

Table 2

ANOVA SUMMARY FOR INTENTION RATINGS OF 20 CONCEPTS
BY 63 RESPONDENTS ON 2 OCCASIONS

Source of Variation	df	Mean Square	Estimate of Variance Component	Proportion of Total Variance
Respondents	62	25.575	0.612	.155
Concepts	19	50.596	0.364	.092
Test Occasions	1	1.625	0	0
Interactions:				
Respondents x Concepts	1178	4.752	1.826	.462
Respondents x Test Occasions	62	1.900	0.040	.010
Concepts x Test Occasions	19	1.756	0.010	.002
Residual	1178	1.101	1.101	.279
Totals	2519		3.953	1.000

TABLE 3
DATA SUMMARY

Source	Product	Intention Measure Procedure		Purchase Measure	Respondent Sample Size
		Scale			
Pickering and Isherwood (1974)	18 generic durable goods (U.K.)	11 point purchase probability (12 month time horizon)	Personal interview	Purchases reported in mail survey conducted 14 months later	276/product (same sample for all 18 products)
Smith, Parker and Davenport (1963)	17 name brand products and services advertised in newspapers	5 point buying plans scale	Personal interview conducted for ad reader-ship study	Purchases reported in telephone interview conducted 11 days later	17 products x 97 respondents/product = 1,649 (data pooled across products)
Twyman (1973)	2 branded packaged goods (U.K.)	7 point purchase intention	Ratings obtained following exposure to commercials	Purchases reported in diaries kept for 1 week following ad test	734 (data pooled for 2 products, same respondents)
Tauber (1975)	3 new packaged goods	5 point purchase intention	Personal interview after trial use in home placement product test	Purchase made over a 4 month period when given home buying opportunity	213/product (same sample for all 3 products)

Table 4

RESULTS FOR JUSTER DATA: LINEAR MODEL

Product	Sample Size	Beta-Binomial Parameters		Linear Model Parameters		$\hat{\rho}$	$\chi^2_{\hat{\rho}}$	R^2	Std. Error of Est., S	R^2_M	Bias, b
		$\hat{\alpha}$ (SE($\hat{\alpha}$))	$\hat{\beta}$ (SE($\hat{\beta}$))	$\hat{A}$ (SE($\hat{A}$))	$\hat{B}$ (SE($\hat{B}$))						
Autos I	$N_1 = 451$	.976	.672	.106	.562	.333	39.394	.772	.069	.109	-.061
	$N_2 = 395$	(.008)	(.077)	(.017)	(.086)						
Autos II	$N_1 = 447$	.109	.537	.071	.528	.387	59.827	.972	.027	.166	.009
	$N_2 = 395$	(.009)	(.050)	(.015)	(.068)						
H/H Appliances I	$N_1 = 2688$	.035	.688	.017	.167	.803	38.587	.890	.009	.069	.023
	$N_2 = 2381$	(.002)	(.046)	(.003)	(.039)						
H/H Appliances II	$N_1 = 2674$	.049	.492	.014	.108	.877	47.540	.752	.019	.091	.066
	$N_2 = 1989$	(.002)	(.025)	(.003)	(.021)						

* Sample size of N_1 is used in the estimation of α and β .Sample size of N_2 is used in the estimation of A and B .

Notes

1. Autos I represent six month intentions vs. six month purchases of automobiles.
2. Autos II represent twelve month intentions vs. six month purchases of automobiles.
3. Appliances I represent twelve month intentions vs. six month purchases of appliances (air conditioners, refrigerators, washing machines, clothes dryers, television sets, and dishwashers).
4. Appliances II represent twenty-four month intentions vs. six month purchases of appliances.

Table 5
OBSERVED AND PREDICTED STATED INTENTION FREQUENCIES
FOR JUSTER DATA

I_x	PRODUCT			
	Autos I	Autos II	H/H Durables I	H/H Durables II
0.0	345 (345)*	293 (292)	2377 (2377)	2174 (2174)
0.1	29 (27)	26 (33)	87 (85)	95 (113)
0.2	16 (15)	21 (20)	57 (46)	99 (63)
0.3	14 (11)	21 (15)	29 (32)	41 (46)
0.4	3 (9)	10 (12)	23 (26)	38 (38)
0.5	10 (8)	9 (11)	22 (22)	28 (33)
0.6	6 (7)	12 (10)	21 (20)	30 (31)
0.7	6 (6)	13 (10)	14 (18)	29 (31)
0.8	5 (6)	11 (11)	11 (18)	36 (33)
0.9	5 (7)	10 (12)	17 (19)	24 (39)
1.0	11 (9)	21 (21)	30 (25)	80 (72)
Computed χ^2_8	6.656	6.937	8.433	32.270

*The figures in parentheses are predicted frequencies rounded to the nearest integer.

Table 6

RESULTS FOR PICKERING AND ISHERWOOD DATA: LINEAR MODEL

Product	Beta-Binomial Parameters		Linear Model Parameters		$\hat{\rho}$	$\chi^2_{\lambda_1}$	R^2	Std. Error of Est., S_e	R^2_M	Bias, b
	$\hat{\alpha}$ (SE($\hat{\alpha}$))	$\hat{\beta}$ (SE($\hat{\beta}$))	$\hat{A}$ (SE($\hat{A}$))	$\hat{B}$ (SE($\hat{B}$))						
Furniture	.096 (.010)	.357 (.040)	.210 (.026)	.303 (.072)	.683	14.164	.685	.070	.043	-.062
Used Cars	.075 (.009)	.484 (.064)	.126 (.021)	.544 (.096)	.426	33.075	.832	.069	.120	-.065
New Cars	.056 (.007)	.389 (.054)	.076 (.017)	.672 (.093)	.298	57.847	.853	.082	.242	-.035
Washing Machines	.097 (.011)	.722 (.095)	.048 (.014)	.439 (.106)	.525	28.999	.560	.109	.164	.018
Record Players	.070 (.009)	.624 (.092)	.066 (.016)	.210 (.095)	.775	7.259	.398	.069	.044	.010
Vacuum Cleaners	.086 (.010)	.637 (.086)	.049 (.014)	.229 (.087)	.754	10.993	.544	.061	.074	.043
Refrigerators	.074 (.010)	.839 (.132)	.047 (.013)	.344 (.126)	.625	14.340	.525	.078	.100	.006
Color Tv's	.049 (.008)	.930 (.183)	.043 (.013)	.316 (.151)	.656	8.642	.440	.065	.071	-.009
Cookers	.068 (.009)	.653 (.099)	.038 (.012)	.233 (.099)	.750	10.664	.659	.045	.087	.034
Black & White TV's	.056 (.008)	.578 (.091)	.044 (.013)	.112 (.078)	.881	3.313	.345	.044	.028	.034

Notes

1. The data in each case represent twelve month intentions vs. fourteen month purchases.
2. Sample size is 276 in each case.
3. Intentions are obtained on eleven point scales ranging from 0 to 10.

Table 7

RESULTS FOR JUSTER DATA: PIECEWISE LINEAR MODEL

Product	Sample Size	Piecewise Linear Model Parameters				R^2	Standard Error of Estimate, S	R^2_M
		Intercept A (SE($\hat{A}$))	Slope 1 $\hat{B}_1$ (SE($\hat{B}_1$))	Threshold $\hat{T}_X^*$ (SE($\hat{T}_X^*$))	Slope 2 $\hat{B}_2$ (SE($\hat{B}_2$))			
Autos I	395	.100 (.016)	.837 (.132)	.500 (.025)	.000 (—)	.921	.041	.123
Autos II	395	.070 (.015)	.640 (.279)	.137 (.273)	.499 (.087)	.974	.026	.166
H/H Appliances I	2381	.017 (.003)	.123 (.177)	.091 (.298)	.177 (.048)	.890	.009	.069
H/H Appliances II	1989	.015 (.003)	.033 (.039)	.219 (.069)	.157 (.034)	.883	.015	.095

Notes

1. Autos I represent six month intentions vs. six month purchases of automobiles.
2. Autos II represent twelve month intentions vs. six month purchases of automobiles.
3. Appliances I represent twelve month intentions vs. six month purchases of appliances (air conditioners, refrigerators, washing machines, clothes dryers, television sets, and dishwashers).
4. Appliances II represent twenty-four month intentions vs. six month purchases of appliances.

Table 8

RESULTS FOR PICKERING AND ISHERWOOD DATA: PIECEWISE LINEAR MODEL

Product	Piecewise Linear Model Parameters				χ^2_2	R^2	Standard Error of Estimate, S	R^2_M
	Intercept $\hat{A}$ (SE($\hat{A}$))	Slope 1 $\hat{B}_1$ (SE($\hat{B}_1$))	Threshold I^*_x (SE($\hat{I}^*_x$))	Slope 2 $\hat{B}_2$ (SE($\hat{B}_2$))				
Furniture	.201 (.026)	.632 (.173)	.300 (.041)	.109 (.115)	1.185	.745	.063	.047
Used Cars	.130 (.021)	.357 (.115)	.600 (.041)	1.068 (.284)	1.543	.876	.059	.126
New Cars	.077 (.017)	.000 (—)	.300 (.023)	1.079 (.121)	5.673	.901	.068	.266
Washing Machines	.049 (.014)	.000 (—)	.200 (.014)	.707 (.157)	4.198	.628	.101	.188
Record Players	.067 (.016)	.000 (—)	.300 (.032)	.421 (.178)	1.659	.490	.064	.055
Vacuum Cleaners	.051 (.014)	.000 (—)	.307 (.120)	.461 (.169)	1.875	.659	.053	.087
Refrigerators	.048 (.013)	.000 (—)	.312 (.092)	.853 (.259)	4.838	.720	.060	.134
Color TV's	.043 (.012)	.000 (—)	.200 (.024)	.561 (.235)	1.932	.500	.061	.086
Cookers	.040 (.012)	.000 (—)	.400 (.044)	.594 (.234)	2.572	.852	.030	.108
Black & White TV's	.043 (.012)	.000 (—)	.500 (.039)	.494 (.269)	2.951	.591	.035	.054

Notes

1. The date in each case represent twelve month intentions vs. fourteen month purchases.
2. Sample size is 276 in each case.
3. Intentions are obtained on eleven point scales ranging from 0 to 10.

Table 9

RESULTS FOR PACKAGED GOODS: LINEAR MODEL

Data Source	Sample Size	Beta-Binomial Parameters		Linear Model Parameters		$\hat{\rho}$	χ^2_1	R^2	Std. Error of Est., S_e	R^2_M	Bias, b
		$\hat{\alpha}$ (SE($\hat{\alpha}$))	$\hat{\beta}$ (SE($\hat{\beta}$))	$\hat{A}$ (SE($\hat{A}$))	$\hat{B}$ (SE($\hat{B}$))						
Smith, et. al. (1963)	1649	.308 (.012)	.474 (.018)	.062 (.007)	.262 (.019)	.687	151.060	.740	.069	.100	.229
Twyman (1973)	734	.586 (.030)	.537 (.027)	.005 (.004)	.201 (.021)	.761	83.324	.540	.100	.163	.412
Tauber (1975) Product A	213	.527 (.053)	.571 (.056)	.000 (—)	.547 (.057)	.303	80.858	.818	.119	.338	.217
Tauber (1975) Product B	213	.834 (.081)	.785 (.074)	.000 (—)	.359 (.040)	.496	40.744	.928	.043	.201	.330
Tauber (1975) Product C	213	1.116 (.110)	.663 (.064)	.000 (—)	.392 (.046)	.434	45.243	.766	.102	.195	.382

Table 10
OBSERVED AND PREDICTED STATED INTENTION FREQUENCIES
FOR PACKAGED GOODS DATA

X	Data Source				
	Smith, et. al. (1963)	Twyman (1973)	Tauber (1975) Product A	Tauber (1975) Product B	Tauber (1975) Product C
0	709 (683)	165 (142)	64 (60)	49 (44)	28 (26)
1	165 (242)	39 (90)	25 (35)	28 (39)	23 (31)
2	99 (192)	66 (79)	26 (31)	36 (39)	45 (37)
3	396 (200)	90 (77)	51 (44)	55 (41)	45 (47)
4	297 (349)	102 (81)	47 (52)	45 (50)	72 (72)
5		119 (97)			
6		153 (168)			
Computed χ^2	269.688	48.677	13.547	9.170	4.059
Critical χ^2	9.210	13.277	9.210	9.210	9.210

*The figures in parentheses are predicted frequencies rounded to the nearest integer.

Table 11

RESULTS FOR PACKAGED GOODS: PIECEWISE LINEAR MODEL

Data Source	Sample Size	Piecewise Linear Model					R^2	Standard Error of Estimate, s	R^2_M
		Intercept $\hat{A}$ (SE($\hat{A}$))	Slope 1 $\hat{B}_1$ (SE($\hat{B}_1$))	Threshold $\hat{I}_x^*$ (SE($\hat{I}_x^*$))	Slope 2 $\hat{B}_2$ (SE($\hat{B}_2$))	χ^2_2			
Smith, et. al. (1963)	1649	.069 (.007)	.135 (.020)	.753 (.021)	1.109 (.117)	39.954	.998	.007	.127
Twyman (1973)	734	.009 (.005)	.007 (.011)	.733 (.018)	1.367 (.134)	67.244	.999	.005	.295
Tauber (1975) Product A	213	.000 (—)	.000 (—)	.399 (.033)	1.122 (.101)	15.473	.960	.056	.403
Tauber (1975) Product B	213	.000 (—)	.000 (—)	.250 (.012)	.553 (.072)	6.052	.996	.010	.231
Tauber (1975) Product C	213	.000 (—)	.000 (—)	.408 (.035)	.816 (.088)	13.225	.962	.041	.252

REFERENCES

Adams, F.G. (1974), "Commentary on McNeil, 'Federal Programs to Measure Consumer Purchase Expectations'," Journal of Consumer Research, 1 (December), 11-12.

Axelrod, J.N. (1968), "Attitude Measures that Predict Purchase," Journal of Advertising Research, 8 (March), 3-17.

Bagozzi, R.P. (1981), "Attitudes, Intentions, and Behavior: A Test of Some Key Hypotheses," Journal of Personality and Social Psychology, 41, in press.

Clancy, K.C. and R. Garsen (1970), "Why Some Scales Predict Better," Journal of Advertising Research, 10 (October), 33-8.

Bird, M. and A.S.C. Ehrenberg (1966), "Intentions-to-buy and Claimed Brand Usage," Operational Research Quarterly, 17 (March), 27-46.

Cox, E.P. (1980), "The Optimal Number of Response Alternatives for a Scale: A Review," Journal of Marketing Research, 17 (November), 407-22.

Cronbach, L.J. (1950), "Further Evidence On Response Sets And Test Design" (1950), Educational and Psychological Measurement, 10 (Winter), 3-31.

_____ and G.C. Gleser, H. Nanda, and N. Rajaratnam (1972), The Dependability of Behavioral Measurements, New York: Wiley.

Davis, H.L. and E.K.E. Ragsdale, "The Newlywed Game Phenomenon: Application to New Product Development," Paper presented at the TIMS/ORSA Meeting, New Orleans, L.A., April 30, 1979.

Dickson, P.R. and W.L. Wilkie (1978), "The Consumption of Household Durables: A Behavioral Review," Report No. 78-117, Marketing Science Institute, Cambridge, MA, December.

Ebel, R.L. (1951), "Estimation of the Reliability of Ratings," Psychometrika, 16 (December), 407-24.

Ferber, R. (1962), "Research on Household Behavior," American Economic Review, 52 (March), 19-63.

Gold, B. and W. Salkind (1974), "What Do 'Top Box' Scores Measure?" Journal of Advertising Research, 14 (April), 19-24.

Gormley, R. (1974), "A Note on Seven Brand Rating Scales and Subsequent Purchases," Journal of the Market Research Society, 16 (July), 242-4.

Granbois, D.H. and J.O. Summers, "Primary and Secondary Validity of Consumer Purchase Probabilities," Journal of Consumer Research 1, (March 1975), 31-8.

Gruber, A. (1970), "Purchase Intent and Purchase Probability," Journal of Advertising Research, 10 (February), 23-7.

Haley, R.I. and P.B. Case (1979), "Testing Thirteen Attitude Scales for Agreement and Brand Discrimination," Journal of Marketing, 43 (Fall), 20-32.

Jackson, D.H. (1967), "Acquiescence Response Styles: Problems of Identification and Control" in I.A Berg, ed., Response Set in Personality Assessment, Chicago: Aldine.

Juster, F.T. (1966), "Consumer Buying Intentions and Purchase Probability: An Experiment in Survey Design," Journal of the American Statistical Association, 61 (September), 658-96.

____ (1974), "Commentary on McNeil, 'Federal Programs to Measure Consumer Purchase Expectations'," Journal of Consumer Research, 1 (December), 12-5.

Kalwani, M.U. (1980), "Maximum Likelihood Estimation of Zero-Order Models Given Variable Numbers of Purchases per Household," Journal of Marketing Research, 17 (November), 549-51.

Keats, J.A. and F.M. Lord (1962), "A Theoretical Model of Mental Test Scores," Psychometrika, 27 (March), 59-72.

Laroche, M. and J.A. Howard (1980), "Non-Linear Relations in a Complex Model of Buyer Behavior," Journal of Consumer Research, 6 (March), 377-88.

Lord, F.M. and M.R. Novick (1968), Statistical Theories of Mental Test Scores, Reading, MA: Addison-Wesley.

McFadden, D. (1970), "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, ed., Frontiers in Econometrics, New York: Academic Press, 105-42.

McNeil, J. (1974), "Federal Programs to Measure Consumer Purchase Expectations, 1946-73: A Post-Mortem," Journal of Consumer Research, 1 (December), 1-10.

Morrison, D.G. (1979), "Purchase Intentions and Purchase Behavior," Journal of Marketing, 43 (Spring), 65-74.

Neter, J. (1970), "Measurement Error In Reports Of Consumer Expenditures," Journal of Marketing Research, V (February), 11-26.

Penny, J.C., I.M. Hunt and W.A. Twyman (1972), "Product Testing Methodology in Relation to Marketing Problems," Journal of the Market Research Society, 14 (January), 1-29.

Pickering, J.F. and B.C. Isherwood (1974), "Purchase Probabilities and Consumer Buying Behavior," Journal of the Market Research Society, 16 (July), 203-26.

Ptacek, C.H. and I. Ross (1979), "Propensity to Buy Ratings," Journal of Advertising Research, 19 (December), 43-7.

Rao, C.R. (1965), Linear Statistical Inference and Its Applications, New York: Wiley.

Rothman, J. (1964), "Formulation of an Index of Propensity to Buy," Journal of Marketing Research, (May), 21-5.

Silk, A.J. (1977), "Test-Retest Correlations and the Reliability of Copy Testing," Journal of Marketing Research, 14 (November), 476-86.

Smith, S., E.B. Parker, and J. Davenport (1963), "Advertising Readership and Buying Plans," Journal of Advertising Research, 3 (January), 25-9.

Snedecor, G.W. and W.G. Cochran (1967), Statistical Methods, 6th ed., Ames: Iowa State University Press.

Sudman, S. and N.M. Bradburn (1974), Response Effects In Surveys, Chicago: Aldine.

Tauber, E.M. (1975), "Predictive Validity in Consumer Research," Journal of Advertising Research, 15 (October), 59-64.

Taylor, J.W., J.J. Houlahan, and A.C. Gabriel (1975), "The Purchase Intention Question in New Product Development," Journal of Marketing, 39 (January), 90-2.

Torgerson, W. (1958), Theory and Method of Scaling, New York: Wiley.

Tversky, A. and D. Kahneman (1974), "Judgment Under Uncertainty: Heuristics and Biases," Science, 185 (September 27), 1124-113.

Twyman, W.A. (1973), "Designing Advertising Research for Marketing Decisions," Journal of the Market Research Society, 15 (April), 77-100.

Warshaw, P.R. (1980), "A New Model for Predicting Behavioral Intentions: An Alternative to Fishbein," Journal of Marketing Research, 17 (May 1980), 153-72.

Wells, W.D. (1961), "The Influence of Yeasaying Response Style," Journal of Advertising Research, (June), 1-12.

_____ (1961), "Measuring Readiness to Buy," Harvard Business Review, 39 (July-August), 81-7.

Wilks, S.S. (1962), Mathematical Statistics, New York: Wiley.

Winer, B.J. (1971), Statistical Principles In Experimental Design, 2nd. ed., New York: McGraw-Hill.

M11-100401



3 9080 004 209 968



BASEMENT
Date Due

JUL 27 1939

Lib-26-67

DARCODE IS ON BACK COVER

